



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# Beryllium Health and Safety Committee Data Reporting Task Force White Paper #2 -- Uses of Uncensored Data

D. H. MacQueen

October 12, 2007

## **Disclaimer**

---

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Beryllium Health and Safety Committee  
Data Reporting Task Force

White Paper #2 – Uses of Uncensored Data

Draft prepared by D.H. MacQueen, Lawrence Livermore National Laboratory  
For the Data Reporting Task Force (DRTF)

Delivered to the DRTF some time after 10/12/07  
May have received subsequent editing by the DRTF

**This draft does not represent the official position of any entity or organization.**

## Introduction

On December 8, 1999, the Department of Energy (DOE) published Title 10 CFR 850 (hereafter referred to as the Rule) to establish a chronic beryllium disease prevention program (CBDPP) to:

- reduce the number of workers currently exposed to beryllium in the course of their work at DOE facilities managed by DOE or its contractors,
- minimize the levels of, and potential for, exposure to beryllium, and
- establish medical surveillance requirements to ensure early detection of the disease.

On January 4, 2001, DOE issued DOE G 440.1-7A, Implementation Guide for use with 10 CFR 850, Chronic Beryllium Disease Prevention Program, to assist line managers in meeting their responsibilities for implementing the CBDPP. That guide describes methods and techniques that DOE considers acceptable in complying with the Rule.

In 2005 a draft DOE Technical Standard “Management of Items and Areas Containing Low Levels of Beryllium” (SAFT 0103; hereafter referred to as the “TS”) was circulated for comment (<http://www.hss.energy.gov/NuclearSafety/techstds/tsdrafts/saft-0103.pdf>). DOE technical standards are voluntary consensus standards developed when industry standards do not exist (see <http://www.hss.energy.gov/NuclearSafety/techstds/index.html> for more information). DOE does not require its field elements to implement DOE technical standards, but field elements may choose to adopt these standards to meet specific needs. This beryllium TS is intended to provide best practices and lessons learned for managing items and areas that contain low levels of beryllium, which has been a costly and technically challenging component of CBDPPs. The TS is also intended to provide guidance for determining if the Rule's housekeeping and release criteria are met. On challenge the TS addressed was the statistical interpretation of data sets with non-detected results, a topic for which no strong consensus exists.

Among the many comments on the draft TS was a suggestion that certain of the statistical comparisons described in the TS could be better implemented if analytical results, even when below a reporting limit, were to be reported by analytical laboratories. See Appendix 1 for a review of terminology related to reporting limits.

The Beryllium Health and Safety Committee (BHSC) formed a Sampling and Analysis Subcommittee (SAS) in 2003. The SAS established a working group on accreditation and reporting limits. By 2006 it had become evident that the issues extended to data reporting as a whole. The SAS proposed to the BHSC the formation of a Data Reporting Task Force (DRTF) to consider issues related to data reporting. The BHSC Board agreed, and requested that the DRTF generate a white paper, to be offered by the BHSC to potential interested parties such as the DOE policy office that is responsible for beryllium health and safety policy. It was noted that additional products could include detailed guidance and potentially a journal article in the future. The SAS proposed that DRTF membership represent the affected disciplines (chemists, industrial hygiene professionals and statisticians, and the DOE office that is responsible for beryllium health and safety policy). The BHSC Board decided that DRTF membership should come from DOE sites, since the focus would be on reporting in the context of the TS and the Rule. The DRTF came into existence in late 2006. The DRTF membership includes industrial hygienists, analytical chemists and laboratory managers, members of the regulatory and oversight community, and environmental statisticians.

A first White Paper, “Summary of Issues and Path Forward”, was reviewed by the BHSC in March 2007 and issued by the DRTF in June 2007. It describes the charter of the DRTF, introduces some basic terminology (reproduced here in Appendix 1), lays out the issues the DRTF is expected to address, and describes a path forward for the DRTF’s work. This first White Paper is available through the BHSC web site.

This White Paper presents recommendations developed by the DRTF following the process laid out in that first White Paper.

### **The basic issue**

The basic issue is exemplified by the following two points of view (see also the first white paper):

- Using all results (including below-RL results) can improve statistical evaluations, and in particular allow release decisions to be made with statistical confidence with fewer samples than if using censored data, and should therefore be an acceptable way of reporting data (i.e., available upon request).
- Reporting uncensored data opens the door to misuse of the reported results, since results at levels below reporting limits do not always represent reliably quantified levels of Be. Noise in the analytical system may dominate the beryllium signal when true beryllium levels in a sample are extremely low. Moreover, such reporting might threaten a laboratory’s accreditation, be inconsistent with accepted technical practices, and create risk communication problems.

### **Discussion**

It was immediately apparent to the DRTF that the purpose of the sampling and the use(s) to which the analytical results would be put, and their regulatory drivers, are crucial factors in determining appropriate data reporting.

The AIHA requires analytical laboratories (that serve the IH community) to establish reporting limits (RLs). RLs are expected to be set so that reportable results (those above the RL) achieve certain goals for precision and accuracy (see the full matrix). Then, any above-RL result is suitable for comparison with a criterion value (i.e., a release limit, a housekeeping limit, an exposure limit, etc.), provided, of course, that the RL is below the limit, and preferably well below the limit.

Since a result indicating levels are above a limit will result in a decision to investigate its cause and potentially take preventative actions, the accuracy and precision of individual measurements near the limit are important. For this reason analytical laboratories are concerned about the potential for incorrect use of uncensored results. In particular, since results below an RL tend to have less precision than above an RL, there is a potential that below RL results may be used for purposes for which their precision and accuracy are not sufficient.

When data are to be used in the aggregate (i.e., a set of several analytical results is to be used together as a whole), then the precision and accuracy requirements for each individual result can be relaxed, because decisions are now made based on properties of the entire set of data. The overall variability of the dataset becomes a key factor affecting decision-making. Even though results below the RL have less precision than results above the RL, a set of data that includes results below the RL is still representative of the population it is intended to represent, and can be used to make (inferential) decisions about that population (provided, of course, that samples were collected from representative locations, but that issue is outside the scope of this White Paper). See Appendix 2 for additional discussion of the statistical motivation underlying this perspective.

However, substantial statistical issues can arise when using uncensored results. In particular, there is the possibility that the distribution of uncensored (below RL) values may be different from the distribution of the above RL values. In this case, using basic parametric statistics that assume the data can be modeled with a single statistical distribution cannot be justified, and reporting below-RL results might not reduce the number of samples needed to support decisions. A second issue is that below-RL values sometimes are negative, in which case the recommended default statistical distribution, the lognormal (Ignacio & Bullock, 2006), can not be used. These are areas for which some work has been done, but additional research (preferably resulting in publication in the peer-reviewed literature) is needed. A third issue is the point at which it becomes worthwhile to use below-RL results. Current parametric statistical methods can handle data sets with 20%, or 30%, or even more, non-detections. It is only when there are too many non-detections that current parametric methods are not viable, but become viable when below-RL values are available. Although recommendations exist regarding how

many is “too many”, the DRTF is not aware of a consensus, and additional research is needed.

The use of below-RL results does not reduce the need to take into account any and all information about the quality of the data. For example, when a laboratory has provided a below-RL number, the value of the RL and its distance from the RL provide information about the quality of the result.

If and when a laboratory agrees to provide uncensored data to a customer, good communication will be necessary. For example, the laboratory’s standard report format may not include a place to report the additional below-RL result, and the laboratory may be reluctant to change the report format. Similar issues may exist with respect to electronic data deliverable formats. Care will be needed to ensure that below-RL results are recognized for what they are; this is particularly important in order to prevent future confusion (i.e., if reports are reviewed years later by people not involved in the original project).

## **Recommendations**

The DRTF examined different drivers and developed a “matrix” of different purposes with a column to indicate when uncensored data might be appropriate for different contexts. The lead author of the matrix was Gary Whitney of Los Alamos National Laboratory. The “matrix” was developed in spreadsheet format, and selected portions are included in this White Paper. The full matrix will be made available at a later date, as part of sampling guidance being developed by the BHSC.

Before presenting our recommendations, we note some basic parameters:

- The DRTF does not under any circumstances support *requiring* the use of uncensored data.
- Uncensored data, if reported, should always be *additional* reporting and at the request of the customer; it should not replace current reporting requirements and should not be used as a standard method of reporting results. Thus, laboratories’ accreditation should not be affected.
- Analytical laboratories, if and when they report levels below their RL, should continue to qualify such results as they do now. Reporting of uncensored data should include appropriate caveats and qualifiers.
- The DRTF anticipates that reporting of below-RL results will be the exception, not the rule.

In addition, the DRTF evaluation was done primarily for ICP-AES and ICP-MS data and for beryllium results, not all analytical methods. The recommendations herein should not be assumed to be applicable to analyses other than these:

- Analytical methods: ICP-MS, ICP-AES, GFAA, AA
- Analytes: Beryllium

The question of whether or not below-RL results might be appropriate to report was examined for each of the following categories:

- Sampling method: Air sampling; Surface sampling; Bulk sampling
- Sampling purpose: Exposure assessment; Control assessment; Housekeeping; Facility assessment; Equipment/material release
- Sampling scope: Breathing zone; Area assessment
- Sampling context: Routine; Event response

Within the scope of these categories and the parameters described above, the DRTF suggests that the use of uncensored data in the situations indicated in the matrix below should be recognized as acceptable. The full matrix should be consulted for more complete information. Any such use should involve appropriate subject matter experts, including a statistician (for dealing with the statistical issues described above), and an analytical chemist (to ensure that the nature and limitations of below-RL results are understood).

## Sampling Matrix

Application	Drivers	Use of Non-Censored Data? <sup>(4)</sup>	Explanation
<b>Air Sampling</b>			
<b>Exposure Assessment</b>			
BZ	10 CFR 850	no	Typically a single sample for the work shift. Sometimes multiple samples for the shift will be used to determine a single TWA.
BZ		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Collective personal sampling data, statistically analyzed to evaluate exposure monitoring results for individual workers (mean, confidence, trends, etc.).
BZ		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Collective sampling data is statistically analyzed to determine if the operation is meeting the criteria or if sampling effort may be reduced. Non-censored data may be useful in evaluating operations where a limited number of samples are available, or where a large number of samples are available, but a very high percentage of data is below the reporting limit. Non-censored data could also be useful for analyzing sub-sets of sample data for the operation (e.g. different workers, different shifts, different time periods, etc.).
BZ		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Representative sampling of selected workers to determine the exposure of a group. Workers to be sampled may be randomly selected or may be chosen based on potential for exposure. Non-censored data may be useful in evaluating operations where a large number of data points are available (different days, multiple workers, long term) and a high percentage of data is below the reporting limit.
Area		no	Only used to estimate exposure in rare cases where a worker cannot wear a sampling pump (e.g. confined space entry). Technically, it is not in compliance with the rule, but may be the only option for an exposure assessment. May be standard sampling pump or high volume.
BZ		no	Typically a single sample for a short sampling period to determine compliance with short term exposure limits. Sometimes a continuous series of short term samples will be used to characterize an operation or determine a single TWA.
	29 CFR 1910.1000		



<b>Air Sampling</b>			
<b>Control Assessment</b>			
Area	10 CFR 850	no	Area samplers may be placed at various locations in the workplace to monitor beryllium levels. High volume samplers are preferred as they provide better sensitivity. Single samples are evaluated to determine if there is detectable beryllium at that location; if criteria have been exceeded; or if a fault or failure of a control is indicated.
Area	Site CBDPP (enforceable under 851)	Yes, when a high percentage of samples are below reporting limits.	Area samplers may be placed at various locations in the workplace to monitor beryllium levels. High volume samplers are preferred as they provide better sensitivity. Samples from multiple locations or multiple samples from single locations may be evaluated to determine if controls are adequate to meet the criteria or if trends or changes are observed in beryllium levels. Non-censored data might be useful if a large percentage of the values are below the reporting limit.
BZ		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Collective personal sampling data statistically analyzed to determine if operation is meeting the criteria. Samples from multiple locations or multiple samples from single locations may be evaluated to determine if controls are adequate to meet the criteria or if trends or changes are observed in beryllium levels. Non-censored data may be useful in evaluating controls for operations where a number of data sets are available (different days, multiple workers, long term, etc.) and a high percentage of data is below the reporting limit.
Area	none	no	Diagnostic sampling to collect data that helps identify sources of airborne beryllium for an operation or determines the level of beryllium being controlled. Not for assessing worker exposure or for comparison to exposure criteria. Sampling could be totally within a control system (e.g. inside a hood, glovebox or enclosure. Typically short term samples.

Surface Sampling			
Housekeeping			
Routine	10 CFR 850	no	Routine sampling to confirm housekeeping efforts and identify specific locations requiring cleaning and/or increased housekeeping efforts. In most cases there would be selective sampling in the vicinity of beryllium operations. Identified locations may be repetitively sampled on a routine schedule. Random sample locations may also be selected.
Routine	Site CBDPP (enforceable under 851)	Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Routine sampling to evaluate housekeeping efforts and identify areas requiring additional housekeeping efforts. In some cases there would be selective sampling in the vicinity of beryllium operations. Random sample locations may be selected within a facility as a part of general long term evaluation of housekeeping efforts.
Event Response		no	Initial sampling in response to a spill or other off-normal event to determine if contamination occurred. Likely to be selective sampling at location of event and random sampling in the general area.
Event Response		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Samples collected in response to a spill or other off-normal event. Sampling data used to determine if beryllium contamination is a concern; to determine extent of contamination; and to evaluate the effectiveness of any clean-up efforts. In most cases there would be selective sampling at the location of a small spill or event to determine if contamination was a concern. Random sampling and statistical analysis performed to evaluate contamination extent and clean-up efforts.

Surface Sampling			
Control Assessment			
Routine	10 CFR 850  Site CBDPP (enforceable under 851)	Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Routine sampling to confirm control efficiency. In most cases there would be judgmental sampling at selected locations. Random sample locations may be selected within facility control systems as a part of general long term evaluation of control efficiency. Control charting or other statistical process control methods may be applied. In this case non-censored data might be useful.
Event Response	Site-specific procedures (not enforceable)	Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	Samples collected in response to control changes, failure or other off-normal event to evaluate the situation as a part of a hazard assessment. In most cases there would be selective sampling at the location of the operation and the control system. In some rare situations there could be random sampling and statistical analysis performed to evaluate large scale control systems such as ventilation systems that support multiple operations. In this case non-censored data might be useful.

Surface Sampling			
<b>Facility Assessment</b>			
Scoping	Site CBDPP (enforceable under 851)	no	This would be the preliminary or investigative sampling performed to determine if there may be beryllium contamination in the facility. Samples typically collected at locations likely to collect and retain contamination.
Scoping	10 CFR 850 (implied)	Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	This is more detailed scoping and is likely to include some selective sampling at locations such as ventilation systems and some random sampling in the general facility. In the case of the random sampling, non-censored data might be useful.
Evaluation	DOE G440.1-7A (guidance)	Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	If scoping indicates contamination or a review shows a history of beryllium operations in the facility, survey units would be identified and additional samples collected to better characterize the situation. This is likely to include random sampling within the defined homogeneous survey units. Non-censored data might be useful.
Clearance/Release		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	This would be random sampling within the survey units of the facility to determine it can be determined to be "clean". If 95-95 UTL is below release limit, then the survey unit is considered clean. In this case non-censored data might be useful.

Surface Sampling			
Release			
Single Item	10 CFR 850Site CBDPP (enforceable under 851)	No if determination is made on any single sample result. Yes if set of samples statistically analyzed and collecting more than 59 samples isn't practicable and a high percentage are below reporting limits.	e.g. Small tool or fixture may have entire surface sampled with one wipe. If contamination above release limit, item would be cleaned and sampled again.e.g. Moderate sized item may have a few surface samples collected. If any individual sample showed contamination above release limit, item would be cleaned and sampled again. Some professional judgment may be required in determining number of samples to collect.e.g. Large, complex item such as a lathe used to machine beryllium. Item would be cleaned and accessible surfaces would be sampled. If 95%-95% UTL is above release limit, item would be cleaned and sampled again.
Batches or "Streams" of Similar Items		Yes, when collecting more than 59 samples isn't practicable and/or a high percentage are below reporting limits.	e.g. Large lot of machine shop tools and fixtures that have gone through a cleaning process. A given percentage would be selected at random and sampled. If 95-95 UTL is below release limit, then the lot is released. Non-censored data may be useful if a large percentage of the data are below the reporting limit.

Bulk Sampling			
Compare beryllium concentrations on surfaces with local soils	10 CFR 850	no  (note: Bulk samples are subject lack of homogeneity and matrix effects during analysis. There may be a lot of "noise" at low levels. The use of censored data is recommended.)	This is the alternate to the 0.2 ug/100 cm <sup>2</sup> release limit in 10 CFR 805.31(b)(1). The contamination on a surface may exceed 0.2 ug/100 cm <sup>2</sup> provided it can be statistically demonstrated that the concentration of beryllium in the bulk material on items to be released does not exceed that of the local soil.
Identify source of beryllium contamination	none	Yes, in cases with more than a slight proportion of data below reporting limits, but not for compliance purposes.	An example would be using a "fingerprint" analysis to look at Be and other metals to determine a likely source. The association of beryllium with other metals (e.g. copper, lead, uranium) may indicate a specific operation as a source. The ratio of beryllium to other naturally occurring metals such as iron or aluminum may indicate whether or not it is consistent with that in the local soils.
Waste Characterization	EPA  10 CFR 850	no	This would be sampling to characterize wastes for proper handling and disposal. Note that only unused beryllium powder is a RCRA waste. However, all beryllium containing wastes must be identified and handled in accordance with the beryllium rule and local and site regulations.
Preliminary Remediation Goals (PRGs)	EPA (region and state)	no	This would be soil sampling to identify and characterize beryllium soil contamination at the site. Also sampling to evaluate remediation efforts.

## Bibliography

American Industrial Hygiene Association (AIHA). *AIHA LQAP Policy Document*, AIHA, 2005

Currie, L.A. “*Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry*”, *Analytical Chemistry* 40, 586-593, 1968

Davis, Charles B. and N. Grams. *When Laboratories Should Not Censor Analytical Data, and Why*. Presented at the U.S. EPA 25<sup>th</sup> Annual Conference on Managing Environmental Quality Systems, April 27, 2006, Austin, TX.

International Union of Pure and Applied Chemistry (IUPAC), *IUPAC Compendium of Analytical Nomenclature, Definitive Rules*, Third Edition, 1998

International Organization for Standardization (ISO), *General requirements for the competence of testing and calibration laboratories*, ISO 17025, 2005

Ignacio, J.S. and W.H. Bullock, *A Strategy for Assessing and Managing Occupational Exposures*, Third Edition. AIHA Press, Fairfax, VA. 1998

US EPA. *40 CFR Part 136, Appendix B, Definition and Procedure for the Determination of the Method Detection Limit*, June 30, 1986

US EPA. *Multi-agency Radiation survey and Site Investigation Manual (MARRSIM)*. EPA-402-R-97-016, U.S. Environmental Protection Agency, Washington, DC. NUREG-1575, December 1997.

US EPA. *Revised Assessment of Detection and Quantitation Approaches*, EPA-821-B-04-005, October 2004

US EPA. *Technical Support Document for the Assessment of Detection and Quantitation Concepts*, EPA 821-R-03-005, February, 2003

US DOE. *10 CFR Part 850, Chronic Beryllium Disease Prevention Program; Final Rule*. December 8, 1999

US DOE. *Implementation Guide for use with 10 CFR 850, Chronic Beryllium Disease Prevention Program*, DOE G 440.1-7A, January 4, 2001

US DOE. *Management of Items and Areas Containing Low Levels of Beryllium*. Draft Technical Guidance supporting 10 CFR 850. 2006

## Appendix 1: Basic Terminology

In this White Paper the terms “detection limit”, “quantitation limit”, and “reporting limit” are generic. They represent any of a large number of different methods and formulas that have been developed in the last 40 years or so to represent various aspects of an analytical method’s ability to detect and measure the presence of a specified analyte in an environmental sample. One widely used detection limit is the U.S. EPA’s “method detection limit”, defined in 40 CFR 136 Appendix B (referred to as the “EPA MDL”). The actual values of these limits, in practice, may be established by individual labs (and thus vary between labs), or may be client-specified.

In this White Paper the term “result” refers to a number, produced by an analytical measurement system, that is intended to estimate the concentration of an analyte in a sample. A result may be above or below any of the types of limits mentioned above, or even negative. Whether or not an analytical method can actually produce a negative number depends on the details of the method. Some do. Negative results are not interpreted to mean that there is a negative amount of the analyte in the sample, but rather that intrinsic analytical variation (including noise in the system) produces a scatter of values, and that scatter may include negative values, depending on the details of the method. For example, the ICP-AES method when used for beryllium can produce negative values.

In this White Paper the term “censoring” refers to the practice of reporting an analytical result as less than a limit (any kind of limit), and not reporting the result, when the result is less than that limit. For example, if a result is 0.03, and the limit being used is 0.05, then a report of “<0.05” (only, without the 0.03) would be considered censoring. We also use the conventional term, “non-detection”, to refer to a result reported as a “less-than” value. The term “censored” is the conventional statistical term for this kind of information. It is commonly used in the fields of failure analysis (engineering) and survival analysis (medicine).

In this White Paper the terms “uncensored data” and “below-RL result” refer to results that are below a laboratory’s normal reporting limit; hence “using uncensored data” refers to receiving such data from a laboratory and using it for some purpose.



## Appendix 2: Statistical motivation

The task force considered the data reporting issue within the larger conceptual framework of making decisions on whether compliance with an administrative health protection limit has been demonstrated. This decision has two components:

- 1) Whether sample analysis results show levels are less than a limit; and
- 2) Whether enough samples have been taken to constrain the probability of unsampled conditions exceeding the limit to an acceptably low level of likelihood.

The first question can be answered for specific conditions as long as reporting limits are less than administrative limits. The second question is answered through use of hypothesis testing statistics applied to all samples collected in a survey. The greater the distance between the sample result and the administrative limit the less likely the population of conditions sampled contains values above the limit. This is the question that reanalysis using below-RL data has the potential to help answer.

In particular, the AIHA recommends (Ignacio & Bullock, 2006) using a 95% upper confidence limit for the 95<sup>th</sup> percentile (95%-95% upper tolerance limit; 95-95 UTL) for this decision. If the 95-95 UTL is less than a decision criterion value, then there is substantial confidence that the probability of unsampled conditions exceeding the limit is acceptably low.

A 95-95 UTL is an estimate of a value in the upper tail of the population. In general, there are two approaches to calculating 95-95 UTLs. The non-parametric approach does not assume that any statistical distribution (e.g., lognormal, normal) can be used to represent the population in question. In this case, in order to estimate a value in the upper tail, enough samples must be collected so that the largest among them can reasonably be expected to be in the upper tail. This requires a fairly large number of samples (59 for the 95-95 UTL, to be specific). The parametric approach, in contrast, assumes a statistical distribution. In this case, it is necessary only to use enough samples to estimate the population parameters (e.g., mean and standard deviation) reasonably well. Then, the shape of the distribution is used to estimate values in the upper tail. Typically, fewer samples are necessary to estimate the population parameters than to ensure that the range of sample results encompasses the extremes of the upper tail (referring to the 95<sup>th</sup> percentile as “extreme”).

Using uncensored data permits the use of parametric methods in cases where using censored data would force the use of nonparametric methods.

Another issue with regard to the use of uncensored data is the fact that below-RL results tend to have larger uncertainty than above-RL results. Currie described the experimental statistics principles that should be applied to establishing reporting limits. Laboratories are expected to adapt these principles to the materials, methods and instruments they use when establishing reporting limits for the sampling matrix sent by their customers. Several variables will affect the laboratories decision on what values should be reported. Initial estimates of the level of precision and accuracy will depend on the number of

replicates of blank and spiked media used. Some variables, such as the presence of interferences, will be unknown at the beginning of a sampling campaign.

Reporting limits tend to be somewhat conservative to compensate for these uncertainties. Therefore, below-RL results that are not very much below reporting limits may still have sufficient precision and accuracy to be useful. This led to the suggestion that reanalysis of below-RL data at the end of a campaign may identify that more results are useful than had been anticipated or that relaxing precision and accuracy goals makes sense in the context of the overall decision making process.

### Appendix 3: Examples of use of uncensored data

To answer the question on whether this potential is likely to prove useful in actual practice, the task force evaluated example data sets. The data set shown in Appendix 4 will be used as an example to illustrate findings. This data set consists of sets of 15 surface wipe samples taken from survey units within a facility with a legacy of past beryllium operations. The boundaries of the survey units were based on the judgment. The decision to collect 15 samples per survey unit was based on an estimate that this would be sufficient to support decisions using parametric statistics. However, only four of nine survey units 2, 5, 6 and 7 were judged to have enough results above reporting limits to justify the assumption of log-normality and the use of parametric statistics.

Figure 1: Percent of Values below Reporting Limits

	Survey Unit								
	1	2	3	4	5	6	7	8	9
# of Samples	15	15	15	15	15	15	15	11	15
# Detected	3	9	2	2	7	5	13	3	3
% Censored	80	40	86.7	86.7	53.3	66.7	13.3	72.7	80

For the four survey units with sufficient results, maximum likelihood estimates (MLE) can be calculated for the 95% upper tolerance limit of the 95<sup>th</sup> percentile (95-95 UTL) and for the percent exceeding the administrative limit of 0.2 micrograms per 100 square centimeters ( $\mu\text{g}/100\text{cm}^2$ ).

Figure 2: MLE for 95-95 UTL ( $\mu\text{g}/100\text{cm}^2$ ) and Percent Exceeding and its 90% Confidence Interval

	Survey Unit			
	2	5	6	7
95-95 UTL	0.020	0.008	0.021	0.368
% Exceeding Limit	0.00002	0	0.00013	2.72
LCL	0	0	0	0.309
UCL	0.263	0.020	5.75	13.4

The MLE of the 95-95 UTL for survey units 2, 5 and 6 supports a conclusion that they are in compliance with the  $0.2 \mu\text{g}/100\text{cm}^2$  administrative limit. The results for survey unit 7 do not support the conclusion that it is in compliance. The upper and lower confidence limit for percent exceeding includes 5%, indicating that it is possible that additional sampling might show that it is in compliance.

For the survey units with fewer than 3 detected results it is difficult to draw any conclusions about the 95<sup>th</sup> percentiles since with only 15 samples  $n/n+1 = 15/16 < 0.95$  so it is more likely than not that the true 95<sup>th</sup> percentile is larger than the largest value in the sample. Without a parametric math model it is impossible to draw confident conclusions about the 95<sup>th</sup> percentile. There are two possible ways to improve this through the use of below-RL data shown in Appendix 4.

The results are thought to represent 3 different conditions: 1) clean surfaces in which wipe samples are not different from field blanks; 2) surfaces with settled dust in which beryllium results are similar to background levels in soil; and 3) surfaces on which legacy contamination is present. The negative numbers are the result of subtraction programmed by the ICP manufacturer to compensate for levels expected in blanks. The value subtracted is not a constant but rather based on trace levels of other metals also in the sample. Adding a constant creates a data set of positive values, with a skewed distribution that is more nearly lognormal than normal. However, the addition affects the logs of small values in the left tail of distribution more than the large values in the right tail, which places the assumption of log normality and models based on that assumption in question. Estimates from the complete data should be accurate enough to be useful, but this has not been verified in publications the task force reviewed and appears to an area that requires further research.

This approach results in the estimates shown in figure 3.

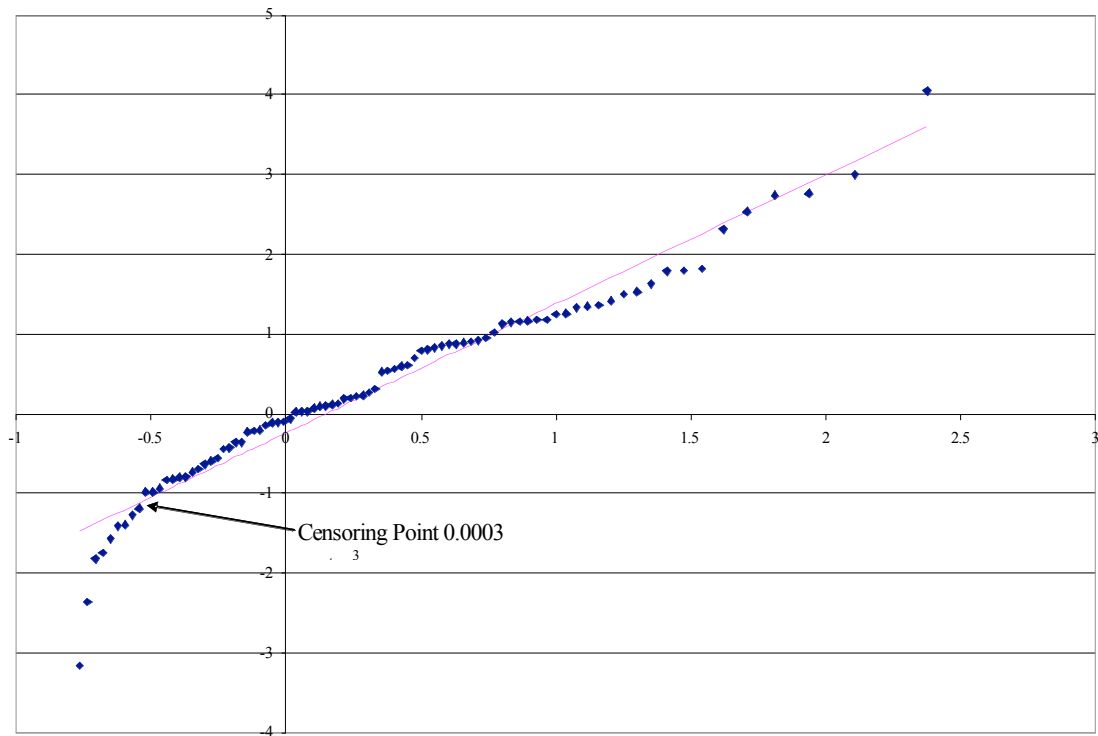
Figure 3: 95-95 UTL ( $\mu\text{g}/100\text{cm}^2$ )

	Survey Unit								
	1	2	3	4	5	6	7	8	9
95-95 UTL	0.01205	0.022642	0.042763	0.008647	0.006502	0.011991	0.25351	0.023544	0.01303

This analysis supports the conclusion that all survey units except 7 are in compliance.

A second approach is to reevaluate the data to establish a censoring point that minimizes censoring to the extent practicable. One approach is to plot the log transformed data to identify the point where the data visibly falls off a fitted line to see if a lower censoring point can be established. The concept is similar to using a calibration curve to establish the reporting limit.

Figure 4: Log Probability Line Fit Plot



The data is reanalyzed using MLE methods for censored data but with 0.0003 instead of 0.002  $\mu\text{g}/100\text{cm}^2$  used as the censoring point. With this censoring point all survey units except 1 and 8 have enough detected results to generate MLE estimates.

Figure 5: Percent Censored and 95-95 UTL ( $\mu\text{g}/100\text{cm}^2$ )

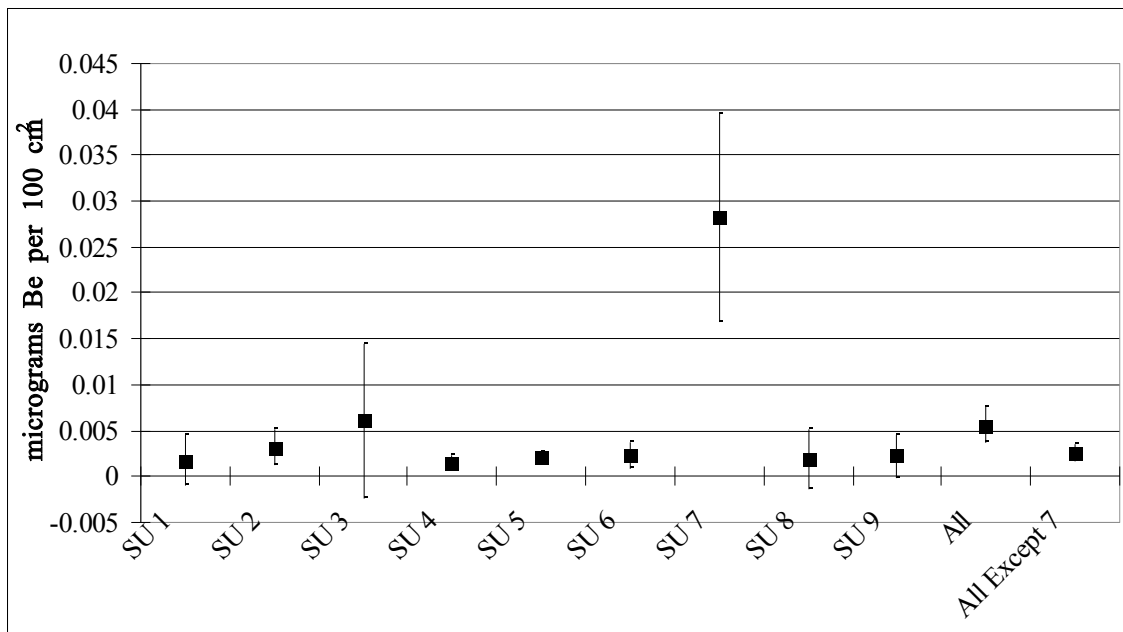
	Survey Unit								
	1	2	3	4	5	6	7	8	9
# Samples	15	15	12	14	15	15	15	11	15
# Detected	4	11	12	11	15	11	15	3	13
% Censored	73.3	26.7	0	21.4	0	26.7	0	72.7	13.3
95-95 UTL	NA	0.069	0.029	0.011	0.009	0.036	0.371	NA	0.013

Figure 5 shows that this approach supports conclusions similar to results arrived at from adding a constant except that survey units 1 and 8 cannot be analyzed. Survey unit 7 is again identified a different from the other in not being clearly in compliance.

A third approach is to revisit the judgments used to draw the boundaries of the survey units in the light of the information now available. If the data supports the judgment that adjacent survey units can be combined, the 59 samples needed to demonstrate compliance with non-parametric methods can be achieved by four survey units. This judgment is supported by estimates of mean and 90% confidence intervals of individual survey units shown in figure 6. The estimates are product limit estimates made from data censored at 0.0003  $\mu\text{g}/100\text{cm}^2$ . Again, survey unit 7 appears to be significantly different

from the others and decisions on whether remediation is needed should be made separately. The others appear similar enough to justify combining with adjacent survey units.

Figure 6: Non-parametric Estimates of Mean and 90% Confidence Interval



In this example, the use of below-RL data aids in the overall decision process by allowing for a more in-depth analysis of subsets of the data to support regrouping. The survey succeeded in identifying a “hot spot” that should be the focus of additional investigation to understand the cause of its difference from other space. Combining results from adjacent clean survey units allows for demonstrating compliance using even the most conservative criteria.

## Appendix 4: Example Data Set

SU	Censored μg/100cm2	Raw μg/SMPL	Raw μg/100cm2	SU	Censored μg/100cm2	Raw μg/SMPL	Raw μg/100cm2	SU	Censored μg/100cm2	Raw μg/SMPL	Raw μg/100cm2
1	< 0.002	-0.00409	0.002	4	< 0.002	-0.00697	0.002	7	< 0.002	0.00301	0.002
1	< 0.002	-0.00318	0.002	4	< 0.002	-0.00697	0.002	7	< 0.002	0.00895	0.002
1	< 0.002	-0.00265	0.002	4	< 0.002	0.00022	0.002	7	0.0057	0.0291	0.0057
1	< 0.002	-0.00224	0.002	4	< 0.002	0.00233	0.002	7	0.0061	0.03114	0.0061
1	< 0.002	-0.00209	0.002	4	< 0.002	0.00287	0.002	7	0.0061	0.03137	0.0061
1	< 0.002	-0.00125	0.002	4	< 0.002	0.0034	0.002	7	0.014	0.07304	0.014
1	< 0.002	0.00049	0.002	4	< 0.002	0.00451	0.002	7	0.016	0.08321	0.016
1	< 0.002	0.00091	0.002	4	< 0.002	0.00462	0.002	7	0.018	0.0929	0.018
1	< 0.002	0.00108	0.002	4	< 0.002	0.00471	0.002	7	0.024	0.12287	0.024
1	< 0.002	0.00129	0.002	4	< 0.002	0.00535	0.002	7	0.033	0.16997	0.033
1	< 0.002	0.00146	0.002	4	< 0.002	0.00573	0.002	7	0.044	0.22783	0.044
1	< 0.002	0.00194	0.002	4	< 0.002	0.01048	0.002	7	0.057	0.29381	0.057
1	0.0034	0.01689	0.0034	4	0.0038	0.01965	0.0038	7	0.057	0.29523	0.057
1	0.004	0.02	0.004	4	0.0062	0.03202	0.0062	7	0.065	0.33414	0.065
1	0.016	0.08229	0.016	5	< 0.002	0.00235	0.002	7	0.073	0.37345	0.073
2	< 0.002	-0.01094	0.002	5	< 0.002	0.00335	0.002	8	< 0.002	-0.01323	0.002
2	< 0.002	-0.00145	0.002	5	< 0.002	0.00422	0.002	8	< 0.002	-0.00939	0.002
2	< 0.002	-0.00143	0.002	5	< 0.002	0.00488	0.002	8	< 0.002	-0.0075	0.002
2	< 0.002	-0.00054	0.002	5	< 0.002	0.00535	0.002	8	< 0.002	-0.00536	0.002
2	< 0.002	0.00298	0.002	5	< 0.002	0.00652	0.002	8	< 0.002	-0.00473	0.002
2	< 0.002	0.00364	0.002	5	< 0.002	0.00937	0.002	8	< 0.002	-0.00425	0.002
2	0.0024	0.01218	0.0024	5	< 0.002	0.00954	0.002	8	< 0.002	-0.00418	0.002
2	0.0024	0.01242	0.0024	5	0.0023	0.01143	0.0023	8	< 0.002	-0.00125	0.002
2	0.0025	0.01264	0.0025	5	0.0026	0.01284	0.0026	8	0.0023	0.01166	0.0023
2	0.0025	0.01309	0.0025	5	0.0027	0.01346	0.0027	8	0.0053	0.02655	0.0053
2	0.0032	0.01646	0.0032	5	0.0034	0.01692	0.0034	8	0.011	0.05268	0.011
2	0.0047	0.02393	0.0047	5	0.0041	0.02031	0.0041	9	< 0.002	-0.00233	0.002
2	0.0061	0.03111	0.0061	5	0.0043	0.02139	0.0043	9	< 0.002	0.00127	0.002
2	0.0061	0.03138	0.0061	5	0.0047	0.02331	0.0047	9	< 0.002	0.00195	0.002
2	0.016	0.08021	0.016	6	< 0.002	-0.00364	0.002	9	< 0.002	0.00203	0.002
3	< 0.002	0.00227	0.002	6	< 0.002	-0.0015	0.002	9	< 0.002	0.00229	0.002
3	< 0.002	0.00363	0.002	6	< 0.002	-0.00025	0.002	9	< 0.002	0.00251	0.002
3	< 0.002	0.00414	0.002	6	< 0.002	0.00084	0.002	9	< 0.002	0.0026	0.002
3	< 0.002	0.00419	0.002	6	< 0.002	0.00158	0.002	9	< 0.002	0.00277	0.002
3	< 0.002	0.00557	0.002	6	< 0.002	0.00468	0.002	9	< 0.002	0.00533	0.002
3	< 0.002	0.0059	0.002	6	< 0.002	0.00583	0.002	9	< 0.002	0.00569	0.002
3	< 0.002	0.00634	0.002	6	< 0.002	0.0065	0.002	9	< 0.002	0.0063	0.002
3	< 0.002	0.0071	0.002	6	< 0.002	0.00875	0.002	9	< 0.002	0.00679	0.002
3	< 0.002	0.00891	0.002	6	< 0.002	0.016	0.002	9	0.0024	0.01189	0.0024
3	< 0.002	0.00914	0.002	6	0.0024	0.01244	0.0024	9	0.0033	0.01634	0.0033
3	0.0035	0.01816	0.0035	6	0.0028	0.0144	0.0028	9	0.021	0.10406	0.021
3	0.033	0.29751	0.033	6	0.0032	0.01669	0.0032				